Comments on

“InundatEd: A Large-scale Flood Risk Modeling System on a Big-data - 1 Discrete Global Grid System Framework”

Chiranjib Chaudhuri, Annie Gray, and Colin Robertson

The paper presents a simple flood modelling framework model based on HAND to predict flood levels in two watersheds in Ontario and Quebec using a big-data discrete global grid systems-based architecture with a web-GIS platform. The authors indicate that the combination of simple conceptual flood method with big-data approaches remains largely uninvestigated, but they don’t make a clear demonstration of what their big-data processing system brings that cannot be accomplished by existing large-scale flood modelling methods at the continental scale.

There is little mention of uncertainty concerning flood estimates in this study, and some of the discrepancies between theoretical and estimated values (e.g. Figure 7b; Figure 8 for large return periods) are dismissed without enough analysis on the implications on the predicted flood zones. The justification for some methodological steps needs to be improved, for example the Lotter method, which is used here despite being singled out in the cited reference (Tullis, 2012) as the only approach “not recommended for use”.

The paper is very long, with several figures, and could be better synthesized to focus on the novelties brought by this study, since there are many other large-scale flood modelling approaches now available. I made suggestions to remove some figures below. I believe that a shorter version, which would include the CSI to better compare with other large-scale flood modelling approaches, could be acceptable for publication once the comments identified below have been addressed.

Detailed comments

Line 71: The list of references for simpler models cited here should include large-scale flood modelling approach such as LISFLOOD-FP that have been used successfully to produce flood maps at the continental scale (e.g. Wing al. 2017). Also, one of the cited references (Oubennaceur et al. 2019) state on p. 46 that “Inundation maps of the Richelieu River were derived with the 2D simulator H2D2” (where H2D2 is a 2D hydrodynamic model). If I understand well their approach, they used this model’s results to develop a simple power function relating discharge to water surface elevation, but it is not clear how they could have obtained this relationship without the H2D2 model. Therefore, can this study be considered a “simple conceptual model”?

Line 171: What is the resolution of the DEM, and what is the vertical accuracy? The LULC data should also be described in more detail (resolution, accuracy). The drainage area of both watersheds (Grand River and Ottawa River) should also be provided here (this information is given in Table 2 which is only presented on line 377).

Line 174: It would be useful to add the (32 and 54) gauging stations used in the analysis for both watersheds on Figure 1. Why is the legend for topography for the Ottawa River starting at a negative value (-71 m)?

Line 198: Is the density of gauging station in Canada comparable to that in Norway in the study of Hailegeorgis and Alfredsen (2017), and does this make a difference in our confidence in a regional approach in Canada? The reference should be Dalrymple (instead of Darlymple).

Lines 215-216: "Only stations with a period of record >= 10 years of annual maximum discharge were maintained (n = 32 and n = 54, respectively)." A minimum of 10 years of annual maximum discharge values seems very low (the minimum in Hailegeorgis and Alfredsen, 2017 was 22 years). It would also be useful to add "for the Grand River and Ottawa River watersheds" before "respectively" as it is not obvious in this sentence.

Line 217: Providing the median or average period of records for both watersheds would be useful.

Lines 246-247 : " for use in a watershed where the flow has been modelled due to human abstraction is a fundamental step of the analysis process and must account for disturbance-related changes to the extreme value characteristics of the flow". It is not clear what you mean by "modelled due to human abstraction" in the context of the two studies watersheds, or what disturbance-related changes are expected, so providing more information here would be useful (some of that information is presented later in Table 2). The Ottawa River, for example, is a very large watershed, with its upstream parts mainly forested, so you need to clarify what disturbances have affected its hydrology since in Table 2 you indicate only 6% farmland and < 2% developed.

Lines 249-250: Here again, it would help to clarify what artificial abstraction you are referring to and how this is supposed to affect extreme value characteristics of the flow. If we look at the causes of the major floods of 2017 and 2019 in the Ottawa watersheds, they look mainly natural (very snowy winter followed by very wet spring, with deeply frozen soil due to very cold temperatures in the autumn, thus limiting infiltration).

Line 259 : " Per assumption c of the index flood method…". More information is needed to understand what assumption c entails. A reference would also be useful.

Line 262 : Qi needs to be defined and Qtilde should be more clearly defined as the median annual maximum discharge.

Lines 298-299: Note that other large-scale flood modelling approaches also don't require channel geometry (e.g. Wing et al. 2017)

Lines 311-312: Do you have an uncertainty estimate on slope estimated with the 30m x 30m DEM? As indicated above, having more information on the vertical accuracy of the DEM would be useful.

Line 320: Why did you choose the Lotter method? In Tullis (2012), it is stated (p. 72) that "Pillai (1962) concluded that the Horton relationship performed the best and that the Lotter relationship gave inconsistent results." On the same page, Tullis (2012) indicates: "Flintham and Carling (1992) evaluated the Horton, Colebatch, Pavlovskii, and Lotter methods. They concluded that the Pavlovskii relationship was the most accurate, the Horton and Colebatch relationships were satisfactory, and the Lotter relationship performed poorly. Four of the five relationships evaluated in the three different studies were identified at least once as a "best performer," but consensus was not achieved regarding an overall best method. The Lotter relationship, on the other hand, was singled out in each study as "not recommended for use."

Line 325: (Figure 4) This figure has a lot of details which may not be needed, particularly since the HAND method is widely known.

Line 397: These 2 stations should be identified on Figure 1

Line 403: divided (instead of divded)

Line 447: Are there references related to flood modelling validation for the use of Matthews Correlation Coefficient? The reference cited here (Chicco & Jurman, 2020) is in the Genomics field. The Critical Success Index (CSI) seems more commonly used for flood modelling validation (e.g. the two listed references), so why not use it here to facilitate comparisons with other studies, for example Wing et al. (2017) who obtained a score of 55.2% for their flood maps of the United States.

Line 456 (Figure 6): I don't think this figure is needed. Providing HAND results (what are the units on Figure 6a,b?) at the scale of such watersheds is not really useful, and there is also little value to showing the drainage network or the Manning's n values, already presented in a table.

Lines 464-466: "The difference in correlation quality can be accounted for in part by the difference in the relative complexities of the delineated networks of the Grand River and Ottawa River watersheds." This is not a sufficient explanation for such a marked difference in correlation values between the two watersheds. What are the "relative complexities" of the network of the Ottawa River watershed that would explain that several points are not at all following the 1:1 slope?

Line 482: The information on dams on the Ottawa and Grand River watersheds should be provided here. To the best of my knowledge, about 40% of the flow is controlled by dams. What is the situation on the Grand River? As indicated below, dam information presented in supplementary material could easily be integrated into Figure 1.

Lines 490-491: "As expected, for the stations with high observation counts (n = 101 and n = 84 for the Grand River watershed (Figure 8a) and Ottawa River watershed (Figure 8b), respectively) the theoretical and estimated return periods are closer, at least for lower return periods." I find this sentence confusing. First, it seems to imply that there are stations with lower observation counts, but these are the only two stations presented. Then, the theoretical and estimated return periods are indeed close for low return period, but not at all for longer return period (even when less than the 5T threshold), which seems problematic.

Lines 497-505: The link between this paragraph and the previous ones is not clear and, overall, this paragraph seems out of place. It is the first time that the hexagonal gridding system is mentioned, and it is difficult to understand why it is problematic. I suggest removing this paragraph.

Lines 518: Why did you not include the CSI, since you are testing 4 metrics?

Line 520: (Figure 10): Why is the scale not the same in each of the figures? And why is the area covered for RP 42.69 different from the other maps? Table 4 should be mentioned here, as it is easier to compare with other large-scale flood modelling approaches (e.g. Wing et al. 2017) with actual values than with maps. It would also be easier to use the CSI for this comparison.

Line 527-528 : Wing et al. (2017) also used a 30-m DEM, and Sampson et al. (2015) a 90-m DEM, and they obtained a fit index of 55.2% and 75%, respectively. Perhaps the problem is more related to the HAND approach compared to the hydraulic modelling approach?

Line 531: Considering that you are not providing a lot of information on the uncertainty in, for example, the slope estimates (see above comment), it remains difficult to be convinced that the problem is with the flood extent polygons. It is also interesting to note that Lim and Brandt (2019), cited here, list CSI in their approaches, but not MCC. Further justification is needed for not using CSI in this study.

Line 545: Again, it is difficult to understand why an index that would allow for comparison with previous studies (such as the CSI) was not used. Providing references where MCC was used to assess the success of simulated floods would be useful.

Line 550: The depth value should still be indicated in the green-red colour scheme on Figure 11. Otherwise, you should use a single fill colour (the same comment applies to Figure 12).

Line 554: It seems obvious that a MCC of 0.95 is associated with a strong model success, so I am not sure to understand what is meant here.

Line 563: The position of dams should be indicated in Figure 1, not in supplementary material.

Line 568: It is possible (instead of it's possible)

Line 576: So a single fill colour should be used instead of the green-red legend in Figure 12.

Line 593: This should have been mentioned earlier, when presenting Figures 9 and 10.

Lines 605-607: “The moderately high FDR value of 0.44 for the 42.69-year return period and the observed overestimation of flood extent (Figure 12B) may be a result of high local Manning’s n values.” It is not clear why high Manning’s n values would only play a role in this case.

Line 611: relatively (instead of realtively)

Line 613: It is the first time the burning of the polygon network is mentioned.

Supplementary material: Table S1: The number of digits after the decimal point should be consistent and reasonable (e.g. 5 digits for discharge values in m^3/s is too many). The same comment applies to Table S2.

Line 724: Dalrymple (instead of Darlymple)